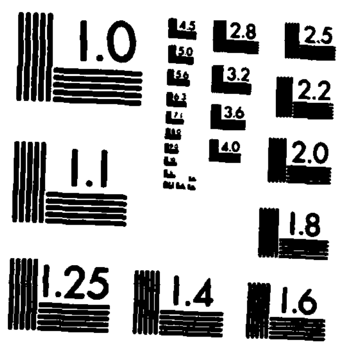
AD-A142 982 THE USE OF DOUBLE SAMPLING IN STUDYING ROBUSTNESS(U) 1/1
PRINCETON UNIV NJ DEPT OF STATISTICS
S MORGENTHALER ET AL. SEP 83 TR-252-SER-2
UNCLASSIFIED ARO-19442.13-MA DAAG29-82-K-0178 F/G 12/1 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

2

AD-A142 982

DTIC FILE COPY

The use of double sampling in studying robustness*

by

Stephan Morgenthaler and John W. Tukey

Technical Report No. 252, Series 2
Department of Statistics
Princeton University
September 1983

*Prepared in connection with research at Princeton University, sponsored by the Army Research Office (Durham). The computing facilities were provided by the Department of Energy. Contract DE AC02-81ER10841.

September 7, 1983

The use of double sampling in studying robustness*

by

Stephan Morgenthaler and John W. Tukey

Technical Report No. 252, Series 2
Department of Statistics
Princeton University
September 1983

Accession For	
NTIS GRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

DTIC COPY INSPECTED

*Prepared in connection with research at Princeton University, sponsored by the Army Research Office (Durham). The computing facilities were provided by the Department of Energy. Contract DE-AC02-81ER10841.

September 7, 1983 84 07 12 040

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(2)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 19442.13-MA	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) Technical Report No. 252, Series 2 "The use of double sampling in studying robustness"		5. TYPE OF REPORT & PERIOD COVERED •
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stephan Morgenthaler and John Tukey		8. CONTRACT OR GRANT NUMBER(s) ~~DE-AC02-81ER10841~~ DAAG29-82-K-0178
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Princeton University Princeton, N. J. 08544		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE September 1983
		13. NUMBER OF PAGES 15
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report deals with an application of double sampling in the area of robustness. Configural polysampling is a technique which allows a detailed comparison of existing estimator and helps in finding smalll-sample-optimal estimators. The technique involves sampling across configurations. The associated sampling error can be reduced by using double sampling. Formulas for doing this are given and demonstrated in an example.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The use of double sampling in studying robustness

Stephan Morgenthaler and John W. Tukey

Technical Report No. 252, Series 2
Department of Statistics
Princeton University
Princeton, NJ 08544

ABSTRACT

This report deals with an application of double sampling in the area of robustness. Configural polysampling is a technique which allows a detailed comparison of existing estimator and helps in finding small-sample-optimal estimators. The technique involves sampling across configurations. The associated sampling error can be reduced by using double sampling. Formulas for doing this are given and demonstrated in an example.

1. Introduction.

Configural sampling (D. Pregibon and J. W. Tukey (1980)) is a powerful tool in studying robust estimators. We want to discuss (in this report) its use in attaining variances and efficiencies (i.e. ratios of variances) for any given location-and-scale-equivariant estimator in various sampling situations. This is obviously an important task in understanding the behavior of estimators across sampling situations and hence in studying robustness. If we are interested in the behavior of any specified location estimator T under any specified sampling situation F, the configural approach works as follows. For a set of configurations $c_1, \ldots, c_N$ drawn at random from situation F, four

Prepared in part in connection with research at Princeton University sponsored by the Army Research Office (Durham). The computing facilities were provided by the Department of Energy, Contract DE-AC02-81ER10841.

September 7, 1983

two-dimensional integrals are calculated (D. Pregibon and J. W. Tukey (1980)). These -- usually numerical -- calculations then allow us to compute the mean-square-error of the estimator T conditioned on the configurations. This conditional mean-square-error will have no sampling error attached to it, its accuracy depends directly on the accuracy of the value of the integrals, which will usually be affected by a numerical error.

The conditional mean-square-errors then have to be averaged across the sampled configurations to get the overall mean-square-error. For a polysampling scheme, configurations are randomly drawn from various situations F, G ... . Then, for each configuration the four integrals are calculated for each situation. In this way, the conditional mean-square-error of T can be calculated in all sampling situations under consideration. Computing weighted means of the conditional m-s-e's across all drawn configurations -- and not just those drawn from a particular situation -- then allows a somewhat more stable overall estimate of the mean-square-errors of T in these situations.

At this second step of the configural approach, i.e. averaging across the sampled configurations, to represent (estimate) the result of averaging over all configurations, a sampling error enters.

This report addresses the question of reducing the sampling error by the method of double sampling (see e.g. Cochran (1977)). In the next section we will give the formulas and in the last section we will discuss an example.

2. Double sampling formulas.

The configural method naturally gives us, in any situation for which we compute the integrals, the (minimal) conditional mean-square-errors and the conditional excess mean-square-error for any location-and-scale-equivariant estimator T. The formulas are as follows: (see D. Pregibon and J. W. Tukey (1980)).

$$m_i^F = \text{minimal cond. mse}_F = \frac{-\text{ave}_F^2(ts^2|c)}{\text{ave}_F(s^2|c)} + \text{ave}_F(t^2s^2|c)$$

$$e_i^F = \text{cond.excess mse}_F(T) = \text{ave}_F(s^2|c)(t_{\text{opt},F} - T(c))^2.$$

Here c denotes the configuration, F the sampling situation (shape or contamponent, for example) and (t,s) are co-ordinates describing the sample y as

$$y = s(t + c).$$

(y and c are n-vectors, s is positive real and t is real. In the last formula it is understood that t is multiplied by an n-vector consisting of 1's).

The polysampling estimate of the overall minimal mean-square-error in situation F is

$$\hat{M}_p^F = \Sigma\, w_i^F\, (\text{min. cond. mse}_F)_i = \Sigma\, w_i^F m_i^F \qquad (2.1)$$

where w_i^F denotes the relative weight of the ith configuration for situation F and m_i^F, as above, stands for

$$\frac{-\text{ave}_F(ts^2|c_i)}{\text{ave}_F(s^2|c_i)} + \text{ave}_F(t^2s^2|c_i),$$

i.e. the minimum mse conditional on the ith configuration. The sums run over the set of all randomly drawn configurations.

The polysampling estimate of the overall excess mean-square-error of the estimator T in situation F is

$$\hat{E}_p^F = \Sigma\, w_i^F\, (\text{cond.excess mse}_F(T))_i = \Sigma\, w_i^F e_i^F \tag{2.2}$$
$$= \Sigma\, w_i^F\, \text{ave}_F(s^2|c_i)\ \ (t_{opt,F,i} - T(c_i))^2$$

where the symbols are as in (2.1).

Double sampling in (2.1) describes the minimal conditional mean-square-errors m_i^F by regression estimates $\hat{m}_i^F$ involving simple functions of the components of the configuration c_i and then gets

$$\hat{M}_p^F = \Sigma\, w_i^F [\hat{m}_i^F + (m_i^F - \hat{m}_i^F)] \tag{2.3}$$
$$= \Sigma\, w_i^F (m_i^F - \hat{m}_i^F) + \Sigma\, w_i^F\, \hat{m}_i^F$$

If we have a regression estimate $\hat{m}^F$ which can be applied to any configuration, we can randomly draw more configurations from the situation F and therefore calculate the second sum in (2.3) with higher accuracy. For these newly drawn configurations we do not have to do the integrations which give the (exact) value m^F. We need only calculate the regression estimate $\hat{m}^F$, which is much simpler and cheaper to do. The double-sampling estimate is therefore

$$\hat{M}_2^F = \Sigma_1\, w_i^F (m_i^F - \hat{m}_i^F) + \frac{1}{N_2} \Sigma_2\, \hat{m}_j \ . \tag{2.4}$$

Here the first sum runs over all configurations where the actual integrals have been computed and the second sum runs over the N_2 configurations drawn for the purpose of double sampling from situation F. In the actual application the number of configurations where the integrals are computed will be small compared to the number of configurations drawn for the purpose of double sampling.*

From (2.4) we can see how double sampling by regression estimates works. The second sum is an estimate of the expected value of the regression estimate in the sampling situation F. The first sum estimates the bias of the regression estimate.

A similar approach to the estimation of the overall excess mean square error (2.2) is now straightforward. Let $\hat{t}_{opt,F}$ be a regression estimate for the cond. optimal location estimate. Then

$$\begin{aligned}\hat{E}_p^F &= \Sigma\, w_i^F \, ave_F(s^2|c_i)\ (t_{opt,F,i} - \hat{t}_{opt,F,i} + \hat{t}_{opt,F,i} - T(c_i))^2 \\ &= \Sigma\, w_i^F\, [ave_F(s^2|c_i)(t_{opt,F,i} - \hat{t}_{opt,F,i})^2 \\ &\quad + \Sigma\, 2ave(s^2|c_i)(t_{opt,F,i} - \hat{t}_{opt,F,i})(\hat{t}_{opt,F,i} - T(c_i)) \\ &\quad + \Sigma\, w_i^F\, ave_F(s^2|c_i)(\hat{t}_{opt,F,i} - T(c_i))^2 .\end{aligned}$$

Here double sampling can be applied in the second sum by introducing a regression estimate for $ave(s^2|\text{configuration})$. This leads to

$$\hat{E}_p^F = \Sigma\, w_i^F [ave(s^2|c_i)\ (t_{opt,F,i} - \hat{t}_{opt,F,i})^2$$

*It should be noted that we have eliminated the use of the relative weights in the second sum by only sampling from situation F. This seems practical and avoids the difficulty of getting the relative weights of newly drawn configurations, which again would involve integration — and maybe another level of double sampling.

$$+ \quad 2\mathrm{ave}(s^2|c_i)(t_{opt,F,i}-\hat{t}_{opt,F,i})(\hat{t}_{opt,F,i}-T(c_i))]$$
$$+ \sum w_i^F(\mathrm{ave}_F(s^2|c_i)-\mathrm{a\partial e}_F(s^2|c_i))(\hat{t}_{opt,F,i}-T(c_i))^2$$
$$+ \sum w_i^F \; \mathrm{a\partial e}_F(s^2|c_i) \; (\hat{t}_{opt,F,i} - T(c_i))^2 \; .$$

In the last sum above only regression estimates occur and we can therefore get a better estimate of this sum by resampling configurations from situation F, which finally yields

$$\hat{E}_2^F = \sum w_i^F[\mathrm{ave}_F(s^2|c_i)(t_{opt,F,i} - \hat{t}_{opt,F,i})^2 \tag{2.5}$$
$$+ 2\mathrm{ave}_F(s^2|c_i)(t_{opt,F,i} - \hat{t}_{opt,F,i})(\hat{t}_{opt,F,i} - T(c_i))$$
$$+ w_i^F(\mathrm{ave}_F(s^2|c_i) - \mathrm{a\partial e}_F(s^2|c_i))(t_{opt,F,\hat{i}} - T(c_i))^2]$$
$$+ \frac{1}{N_2}\sum_2 \mathrm{a\partial e}(s^2|c_j)(\hat{t}_{opt,F,j} - T(c_j))^2,$$

the double-sampling estimate of the overall excess mean-square-error of the estimator T in situation F .

Equations (2.4) and (2.5) give estimates based on the technique of double sampling for quantities we are interested in. An estimate of the efficiency of the estimator T in situation F can be obtained by

$$\hat{\mathrm{eff}}_F(T) = \frac{\hat{M}_2^F}{\hat{M}_2^F + \hat{E}_2^F}$$

which will be a more stable estimate than

$$\hat{\mathrm{eff}}_F(T) = \frac{\hat{M}_p^F}{\hat{M}_p^F + \hat{E}_p^F} \, ,$$

the increase in stability is, however, determined by the number N_2 of configurations in the second sample and -- more importantly -- by the quality of the regression estimates for the three quantities

$t_{opt,F}$

$ave_F(s^2|\text{configuration})$

and

minimal conditional mean-square-error in situaton F .

The final section gives an example of the use of this technique and discusses the problem of getting the regression estimates in a special case.

3. Example.

In order to study robustness properties of various estimators, and in order to define new -- in a small-sample sense optimal -- location estimators, four increasingly heavy tailed shapes -- joining the Gaussian to a Cauchy-like -- are considered in the following experiment. We will call these shapes gupa-nm (Gaussian-Pareto distributions), where n and m are integers such that the tail behavior of the corresponding cumulative distribution function is Paretian with exponent $-(n/m)$. These distributions are such that the central part is exactly Gaussian. The gupa-nm shapes are discussed in Garfinkle (1982). We chose the four shapes gupa60 (i.e. Gaussian), gupa62, gupa64 and gupa66. The diversity of the last one has tail behavior like $(\frac{x-\mu}{\sigma})^{-2}$, and is therefore like a Cauchy density in the tails. For each of these four situations we draw at random 200 configurations, i.e. a total of 800 configurations, for the

case of samples of size 5. This is our primary set of configurations and for each we calculate all of the necessary two-dimensional integrals for all of the four situations. This is a total of 4x4 = 16 integral values for each configuration. Now we are ready to do the configural polysampling. We can estimate for each of the four situations the polysampling estimate $\hat{M}_p^F$ of the minimal attainable mean-square-error (Pitman (1938)). The results are given in Table 3.1.

Table 3.1

Polysampling and single sampling estimates of the
Pitman variances for samples of size 5
(standard errors in parenthesis)

	single sampling	polysampling
gupa 66	.3705 (0.1175)	.3543 (0.1076)
gupa 64	.2744 (0.0465)	.2755 (0.0497)
gupa 62	.2033 (0.0393)	.2065 (0.0287)
Gaussian	.2000 (0.0000)	.2000 (0.0000)

The numbers in parenthesis are estimates of the standard deviations of the estimate. Single sampling refers to the estimate one gets by using only the configurations drawn from the "right" situation. For the Gaussian case there is no error since the integrations can be done analytically, and, since all configurations behave exactly the same way, the sampling error is eliminated. For the gupa66 case we are in slight trouble and it seems worthwhile to apply double sampling. For our primary set of configurations we have 200 gupa66 drawn ones. These we plan to use in order to get the necessary regression estimates. Configurations are, in the case we discuss here, ordered 5-vectors and we choose a normalization such that the second component is fixed at -1 and

the fourth component at +1. We therefore only need to consider the first, c_1, third, c_3, and fifth, c_5, components. We therefore look for regression functions

$$\hat{t}_{opt,\ gupa66}\ (c_1,c_3,c_5),$$

$$\hat{ave}_{gupa66}\ (s^2|\text{configuration})\,(c_1,c_3,c_5)$$

and

$$\text{cond.min.m}\hat{s}\text{e}_{gupa66}\ (c_1,c_3,c_5).$$

We have 200 sets of (c_1, c_3, c_5) - values with the corresponding (numerically computed) responses. This seems a straightforward regression problem. First aid (Mosteller and Tukey (1977)) tells us to use

$$\tilde{x}_1 = \log(-1-c_1)$$
$$\tilde{x}_3 = \log((1-c_3)/(1+c_3))$$
$$\tilde{x}_5 = \log(c_5-1)$$

as our carriers, but this, as trial teaches us, would have the effect of treating the values $c_1 = -1$, $c_3 = \pm 1$ and $c_5 = +1$ in a too extreme way. We therefore propose the use of

$$\begin{aligned} x_1 &= \log(1-\delta-c_1) \\ x_3 &= \log\left(\frac{1-c_3+\delta}{1+c_3+\delta}\right) \qquad (3.1) \\ x_5 &= \log(c_5-1+\delta) \end{aligned}$$

where δ is a small value at our disposal. After a few trials, we choose $\delta = 0.1$.

First aid for the three response variables tells us to use the logarithm for ave(s^2|configuration) and min. cond. mean-square-error, which both only take on positive values. Linear regression can be applied to these re-expressed variables, which we write as

$$u = \log \text{ave}(s^2 | \text{configuration})$$
$$v = \log \text{min.cond. mse} .$$

At this point we need to consider the behavior of x_1, x_3, x_5 and our 3 response variables under reflection of the configuration. This is mostly simply put as

$x_1 + x_5$	--->	$x_1 + x_5$	(even)
$x_1 - x_5$	--->	$-(x_1 - x_5)$	(odd)
x_3	--->	$-x_3$	(odd)
$\hat{t}$	--->	$-\hat{t}$	(odd)
$\hat{u}$	--->	$\hat{u}$	(even)
$\hat{v}$	--->	$\hat{v}$	(even)

Accordingly we should initially approximate $\hat{t}$ by a linear combination of x_1-x_5 and x_3, but $\hat{u}$ and $\hat{v}$ by linear functions of x_1+x_5 above.

We find the following fitted equations

$$\hat{t} = .131\ (x_1-x_5) + .314\ x_3 \qquad R^2 = .96$$
$$\hat{u} = .051 - .280(x_1+x_5) \qquad R^2 = .41$$
$$\hat{v} = -.922 - .131(x_1+x_5) \qquad R^2 = .86$$

The R^2 values for u and v are encouraging, they are, however, not as large as we should like.

We might be able to do better with polynomials in x_1+x_5, x_1-x_5, and x_3 of higher order. For $\hat{t}$ we want expressions that are odd under reflection. Where the simplest possibilities are (a) odd powers of odd expressions, such as

$$x_1-x_5,\ x_3,\ (x_1-x_5)^3,\ \text{and } x_3^3$$

and (b) products of even expressions and odd ones, such as

$$(x_1+x_5)(x_1-x_5) = x_1^2-x_5^2,\ \ x_3^2(x_1-x_5),\ \text{and } (x_1-x_5)^2\ x_3$$

For $\hat{u}$ and $\hat{v}$ we want terms that are even in reflection. Here the simplest possibilities are (a) even expressions and b) squares and cross-products of odd expressions, such as

$$x_1+x_5,\ \ (x_1-x_5)^2,\ \ x_3^2\ ,(x_1-x_5)x_3$$

as well as products and powers of these quantities, such as

$$(x_1+x_5)^2\ ,(x_1+x_5)(x_1-x_5)^2\ ,(x_1+x_5)x_3^2,\ (x_1+x_5)(x_1-x_5)x_3, \text{and } (x_1-x_5)^4$$

Using some of these terms, selected step-by-step on the basis of examining suitable residual plots, leads to fits with multiple-R^2 values above 90%. In the example, this process produced:

$$\hat{t} = .133(x_1-x_5) + .388x_3-.018x_3^2 \qquad R^2 = 0.97$$

$$\hat{u} = -.3685 - .215(x_1+x_5) + .210x_3^2$$
$$+ .159\ x_3(x_1-x_5)-.038\ x_3^2(x_1+x_5) \qquad R^2 = 0.90$$

$$\hat{v} = -.91 -.13\ (x_1+x_5)-.033\ x_3\ (x_1-x_5) \qquad R^2 = 0.95$$

where x_1, x_3 and x_5 are defined in (3.1).

Remark:

If we could also fit the relative weights $w_{gupa66}(c_1, c_3, c_5)$, which would also start as logarithms, we could go ahead and use all the approximations to get an approximation to the bi-effective Gaussian-gupa66 location estimate (Bell and Morgenthaler (1981)). This can indeed be done.

With the above regression estimates we are now ready to compute double sampling estimates according to (2.4) and (2.5). The following table contains the results.

Table 3.2

Double sampling estimates of the Pitman variance for samples for size 5 for the gupa66 situation

	gupa 66
$N_2 = 2000$	.3459
$N_2 = 2000$	.3444
$N_2 = 2000$	.3472
$N_2 = 2000$	.3425
combined	.3450

Each of the above estimates is based on a secondary sample of gupa66 drawn configurations of size $N_2 = 2000$. For each of these configurations we simply have to calculate the regression function and do not have to compute any integrals. We therefore can easily afford to choose the secondary set at least ten times as large as the primary set.

The estimates in table 3.2 are not the more complex ones given in (2.4). Instead of using a polysampling scheme in the first — bias — part of (2.4), they simply apply simple configural sampling throughout

and are therefore of the form

$$\hat{M}_2^F = \frac{1}{N_1} \Sigma_1 \ (m_i^F - \hat{m}_i^F) + \frac{1}{N_2} \Sigma_2 \ \hat{m}_j^F \ , \tag{3.2}$$

Where the notation is as in (2.4), but here the first sum runs only over the gupa66 - drawn configurations in the primary set. In our example we have $N_1 = 200$.

By doing double sampling we got an answer quite close to — and even below — the polysampling answer. The values in table 3.2 are remarkably stable, with a standard deviation of 0.002, but these values are of course correlated. It seems, however, that double sampling gives us an additional decimal place. The following table shows the standard errors of the estimate in table 3.2 depending on the population value p_F of the correlation in our regression function. The formula is (see Cochran (1977), section 12.6, p. 338)

$$\begin{aligned} \hat{\text{var}}(\hat{M}_2^F) &= (1-p_F^2) \ \hat{\text{var}}(\text{single sampling estimate}) \\ &\quad + p_F^2 \frac{N_1}{N_2} \ \hat{\text{var}}(\text{single sampling estimate}) \\ &= (1-(1-\frac{N_1}{N_2})p_F^2) \ \hat{\text{var}}(\text{single sampling estimate}) \end{aligned}$$

(($\hat{M}_2^F$ as in (3.2)).

Table 3.3

Standard errors for the double sampling estimates in the gupa 66 situation

ρ_F^2	$N_2 = \infty$ *	$N_2 = 8000$	$N_2 = 2000$	$N_2 = 200$	$N_2 = 0$ (polysampling)
0.0	.1175	.1175	.1175	.1175	.1076
0.4	.0910	.0918	.0940	.1175	.1076
0.8	.0525	.0551	.0622	.1175	.1076
0.9	.0372	.0411	.0512	.1175	.1076
0.95	.0263	.0319	.0447	.1175	.1076
0.99	.0118	.0219	.0388	.1175	.1076
.992	.0105	.0213	.0385	.1175	.1076
.995	.0083	.0203	.0380	.1175	.1076
.999	.0037	.0189	.0373	.1175	.1076
(1.000)	(0)	(.0186)	(.0372)	(.1175)	(.1076)

(*) This is the contribution from the simple configural estimate of regression function bias.

The estimates of ρ_F^2 we get from fitting the equations, i.e. our R^2 values, are somewhat optimistic. We also have to be careful and transform them to R^2 values for the original — not the re-expressed — response variables. For the variable exp(v) (= minimal conditional mean-square-error) the observed value is $R^2 = 0.95$. Table 3.3 indicates that the double sampling estimates based on $N_2 = 2000$ have about halved the standard error.

It is clear from the table above that, to get a sizable reduction of the sampling error, we must achieve a high correlation. Doing better than $R^2 = .95$ could be quite rewarding, particularly for appropriately

large N_2.

The application of double sampling to the problem of estimating excess mean-square-errors has not yet been undertaken, but we expect about the same reductions.

September 7, 1983

REFERENCES

Cochran, W. G. (1977). Sampling Techniques, John Wiley, New York.

Bell Krystinik, K. and Morgenthaler, S. (1981). Comparison of the bioptimal curve with curves for two robust estimates, Technical Report No. 195, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey.

Garfinkle, S. E. (1982). Junior Paper, Department of Statistics, Princeton University, Princeton, New Jersey.

Mosteller, F. and Tukey, J. W. (1977). Data Analysis and Regression: A Second Course in Statistics Addison-Wesley Publishing Company, Reading, Massachusetts.

Pregibon, D. and Tukey, J. W. (1981). Assessing the behavior of robust estimates of location in small samples: Introduction to configural polysampling, Technical Report No. 185, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey.

END

FILMED

8

DTIC